Chemical Senses Vol.13 no.2 pp.295–305, 1988

A multidimensional statistical study of similarities between 74 notes used in perfumery

Maurice Chastrette, Allal Elmouaffek and Philippe Sauvegrain[1]

Laboratoire de Chimie Organique Physique, Université Lyon 1, 43 Bd du 11 novembre 1918, 69622 Villeurbanne, France and [1]Firmenich SA, Case Postale 239, CH-1211, Genève 8, Switzerland

Abstract. From a data bank of 2467 odoriferous products, the similarities between 74 notes used in perfumery were calculated. The similarity matrix (74,74) contains ~63% of zero values and shows that only a few pairs of notes present high similarity coefficients. A fine analysis using ascending hierarchical taxonomy with the complete linkage procedure shows that 14 notes are isolated while 60 notes are regrouped in 27 groups containing two to four notes. The isolated notes correspond to well-defined structural particularities. Some pairs of notes or groups present similarities in their chemical structures but some groups are built on the basis of semantic processes. This study shows that the notes are generally independent, with no strict hierarchy among them, and rules out the existence of primary odors.

Introduction

The accurate description of odors is of prime importance for communications in the world of perfumery. Dravnieks (1982) recognized two principal methods for this purpose. In reference odorant methods, the odor is described by rating similarities by direct comparison to a series of reference odorants. In semantic methods, the odor is described by words or rated for the applicability of various odor descriptors.

The semantic method is widely used by perfumers, who describe the odors by a relatively small number of words or notes, such as fruity, woody, green. An important point in using this system is to ensure a minimum consensus between different perfumers in the description of odors.

In a recent study, Brud (1986) asked perfumers to associate, with each one of the odor descriptions of a list of 20, the name of only one substance (essential oil, flavour, chemical substance) representing that odour particularly well. The 120 respondents associated 507 substances (pure compounds or mixtures) with the 20 odors, among which nearly 400 were cited by only one person. The consensus was very good for some odors like amber or musk and very poor for others like fatty or floral ones.

It is to be expected that when experts are allowed to meet and to reach some agreement, the correspondence between descriptors and odors will be more satisfying. Harper (1975) summarizes the findings of seven experts who discussed the question of knowing which chemical substances of high purity best represent particular odor qualities. For a list of 44 separate odor qualities, 127 pure substances were cited among which five substances were cited as corresponding to floral and 14 others as corresponding to fruity which had to be split into two different notes.

Dravnieks (1982) reports on the extent to which the semantic profiles of 10 odorants stabilize if a large number of subjects are utilized. He used a sample of 150 subjects with a list of 146 descriptors and found that 'profiles based on combined responses of many subjects are stable constructs'.

If the set of notes used in perfumery is considered as providing an adequate description, the next problem to be solved is to understand the different types of relations between these notes. The need to compare odors led long ago to the construction of numerous classifications, none of which has yet gained wide acceptance. The number of classes varies from four for Crocker and Henderson (1927) to 45 for Cerbelaud (1951) [for reviews about different classifiations, see Kastner (1973), Boelens (1974, 1983), Jaubert (1983), Dore *et al.* (1984) and Jaubert *et al.* (1986)].

Boelens (1983) presents a summary of the principal classifications, limited to the 10 most usual notes, which shows that there is no consensus even on these 10 notes. In fact, there is no consensus even on the principles of classification since the notes have been regrouped either according to their vegetable or animal origin, or according to the resemblances between their chemical structures, or both. In addition, the hierarchy established between the classes and sub-classes owes much to empiricism.

The present work is an attempt to measure the extent to which the most frequently encountered notes are associated in the description of odors in perfumery. Associations among notes should allow us to determine any degree of similarity between notes and eventually between groups of notes, and perhaps, going further, to define more accurately the olfactory universe of perfumers. Moreover, this knowledge may be useful in the establishment of structure−odor relationships. A preliminary note has been published earlier (Chastrette *et al.*, 1986a).

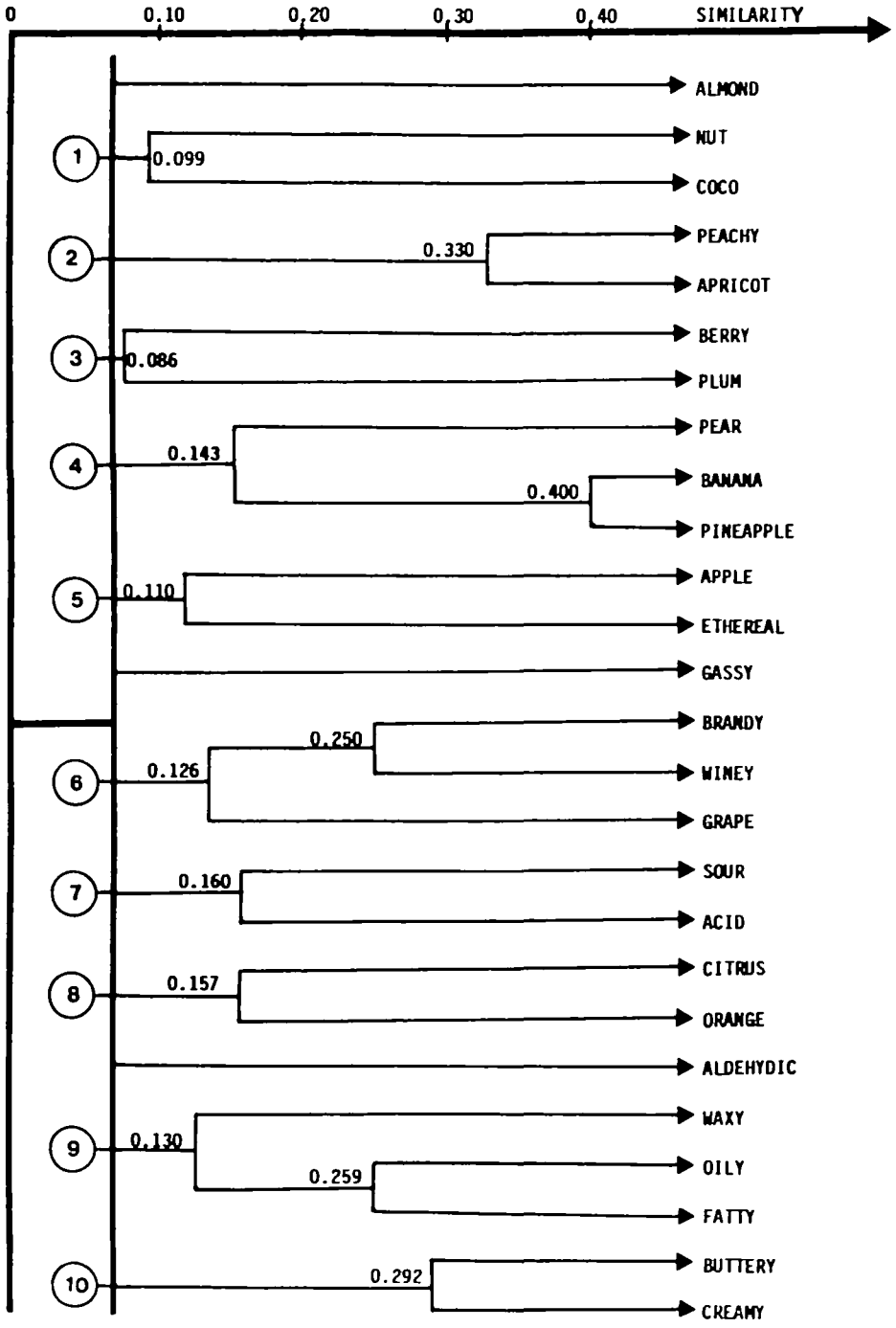
We used a data bank compiled from the reference book of Arctander (1969). It must be noted that Arctander's descriptions are those of one perfumer only, although a very influential and widely recognized one. Similarity coefficients among the 74 most frequently encountered notes in perfumery have been calculated. From these similarity coefficients several hierarchical taxonomies have been effected in order to give a better description of the proximities that exist among notes.

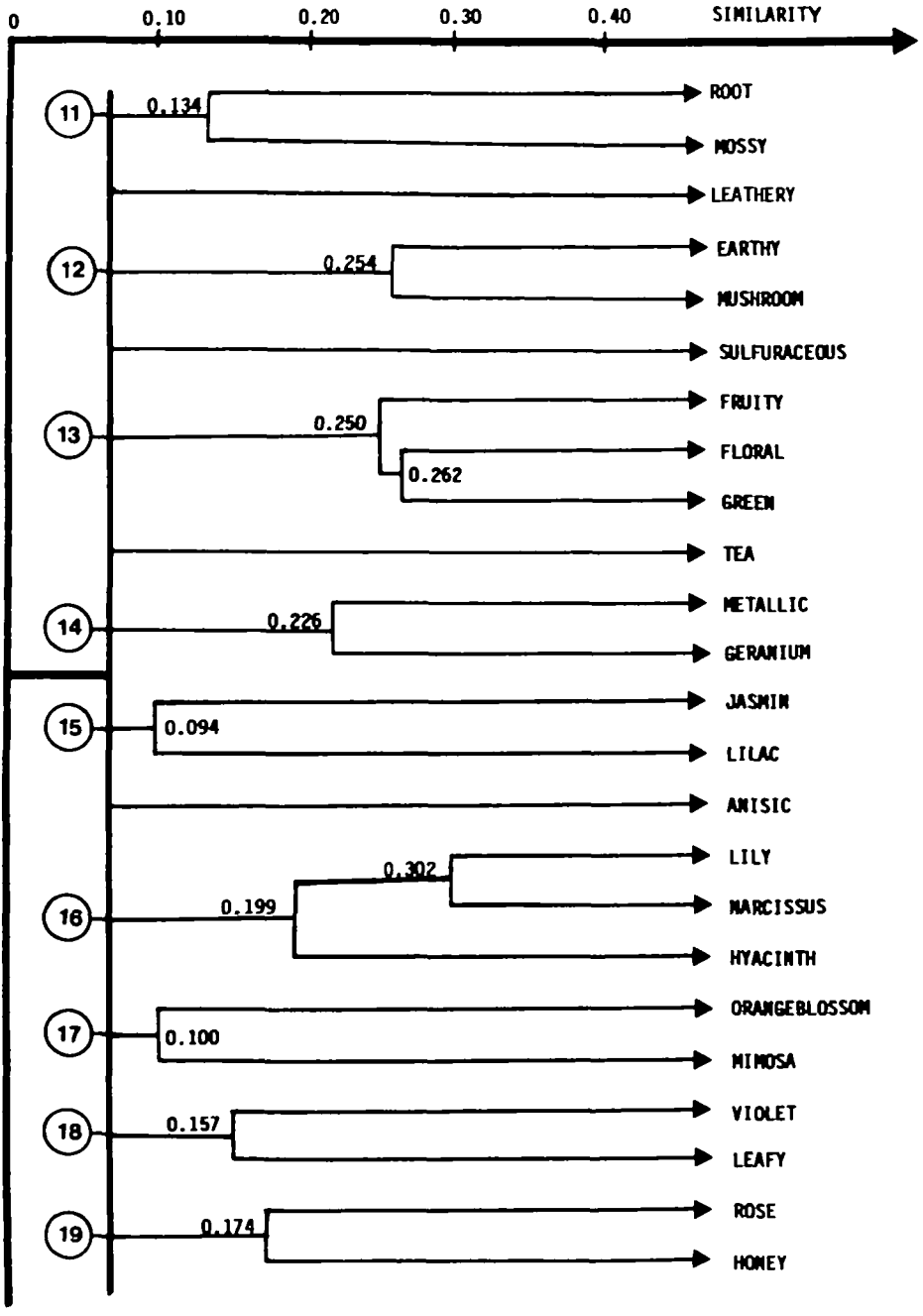
Materials and methods

Selection of olfactory notes

From the reference work by Arctander (1969) we compiled a data bank of 2467 pure substances, the odors of which are described using 270 words. Words such as dry, fresh, strong, weak, warm, chemical, deep, etc., which do not provide qualitative olfactory information, were eliminated. That left 233 notes cited a total of 6768 times, with an average occurrence of 29. Thus the average number of words used for description of the odor of one particular compound is 2.7. However, the 34 most frequent notes represent 80% of the citations.

We selected the relevant notes on the basis of their occurrences, the least frequent notes being useless in a statistical study such as ours. Among the 233 notes, 158 had an occurrence of ≤ 11 and represented only 8.5% of the total citations. Some words, such as amber/ambergris, citrus/lemon, raspberry/berry, were then regrouped, and we removed all notes presenting an occurrence of < 12 after regroupment, except smokey (11 occurrences) and sandalwood (10 occurrences) because of our current interest in these notes. The 74 notes left after this treatment are shown in Figure 1.


0
0,10
0,20
0,30
0,40
SIMILARITY
ALMOND
NUT
COCO
PEACHY
APRICOT
BERRY
PLUM
PEAR
BANANA
PINEAPPLE
APPLE
ETHEREAL
GASSY
BRANDY
WINEY
GRAPE
SOUR
ACID
CITRUS
ORANGE
ALDEHYDIC
WAXY
OILY
FATTY
BUTTERY
CREAMY
1
2
3
4
5
6
7
8
9
10
0.099
0.330
0.086
0.143
0.400
0.110
0.250
0.126
0.160
0.157
0.130
0.259
0.292


0
0.10
0.20
0.30
0.40
SIMILARITY
11
0.134
ROOT
MOSSY
LEATHERY
12
0.254
EARTHY
MUSHROOM
SULFURACEOUS
13
0.250
FRUITY
0.262
FLORAL
GREEN
TEA
14
0.226
METALLIC
GERANIUM
15
0.094
JASMIN
LILAC
ANISIC
16
0.199
0.302
LILY
MARCISSUS
HYACINTH
17
0.100
ORANGEBLOSSOM
MIMOSA
18
0.157
VIOLET
LEAFY
19
0.174
ROSE
HONEY

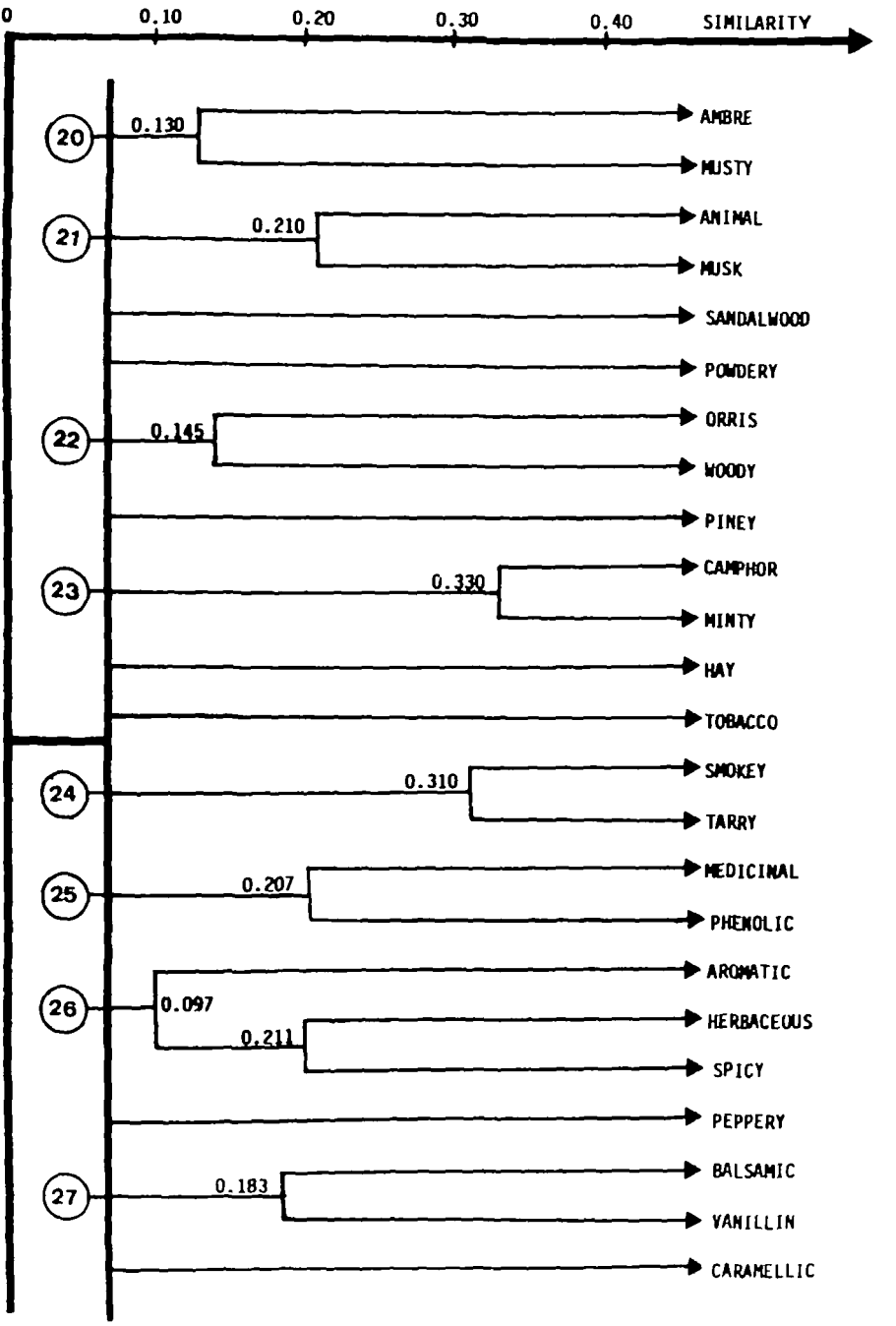


Fig. 1. Dendrogram of 74 notes using the complete linkage method.

Methods

The information contained in the data bank was analyzed in two successive stages: calculation of similarities among notes and then visualization of similarities with the aid of hierarchical classification methods.

We first considered the 2467 substances of our data bank to be individual entities defined by the value 1 or 0 taken by each of the selected 74 variables associated with notes. For example the musk variables takes the value 1 if the compound is musky and 0 in the opposite case. However, from the point of view of the similarity among notes it is more convenient to consider that we have 74 notes defined by their frequency of occurrence in the description of 2467 substances. The matrix of data is thus a matrix (2467, 74) composed of terms taking the values 1 if the note is present in the description of the product and 0 when the note is absent, which is actually more frequently the case. However, classical methods of factorial analysis do not work well in matrices like this one (Iglesias, 1975; Randebrock, 1985; Chastrette *et al.*, 1986a).

In fact it is more convenient to use the square matrix (74,74), diagonal terms of which represent the occurrences of notes, and non-diagonal terms represent the co-occurrences between notes. From this matrix of occurrence/co-occurrence, we calculated the similarity s_{ij} between two i and j notes. Among various similarity coefficients proposed in the literature (Cole, 1949; Sokal and Sneath, 1963; Iglesias, 1975; Benzecri, 1980) we selected the one proposed by Ochai (1957) which does not take into account the case where the substances present neither the i note nor the j note. The s_{ij} coefficient is defined by:

$$s_{ij}=a_{ij}/[a_{ij} + a_j)(a_{ij} + a_j)]^{1/2}$$

a_{ij} : number of products presenting both i and j notes

a_i : number of products presenting the i note only

a_j : number of products presenting the j note only

The s_{ij} coefficients are the terms of the similarity matrix, S (74,74). A broad analysis of similarities was done directly on the similarity matrix.

A finer analysis was then achieved using two methods of hierarchical agglomerative classification, the average linkage method and the complete linkage method (Massart and Kaufman, 1983; Dagnelie, 1984; Roux, 1985). The distance between two i and j notes is taken as $1-s_{ij}$. In the complete linkage the distance between two clusters is considered to be equal to the largest distance between two individual notes, one of each cluster. In the average linkage, when two clusters A and B have been joined, the distance between the new cluster and an already existing cluster C is calculated as the average distance between each of the two clusters A and B and the cluster C. In the presentation of our results the complete linkage method will be preferred but any eventual divergence between the two methods will be pointed out.

Results and discussion

The similarity matrix, S, contains the coefficients of Ochiai for the 2701 possible pairs of notes.

We consider that two notes have a negligible similarity when the coefficient of Ochiai is lower than 0.08. This threshold value is observed when i and j notes have an 8% co-occurrence in the case of an approximately equal number of occurrences (for example eight co-occurrences for occurrences of i and j notes equal to 100). With different numbers of occurrences, e.g. 24 for the i note and 100 for the j note, this limit corresponds to four co-occurrences (17% of i notes and 4% of j notes).

The most striking result is the generally weak similarity of notes: 1691 (62.6%) out of the 2701 similarity coefficients have a zero value, corresponding to no co-occurrence between the corresponding pairs. Only 224 similarity coefficients (8.3% of the total) are >0.08. It is worth noting that out of these 224 associations, 110 are brought along by only five notes (28 by fruity, 23 by floral, 21 by herbaceous, 20 by green and 18 by woody). The 69 other notes give only 114 associations, with an average of 1.65 associations per note. Moreover, very few similarity coefficients have a value >0.3. The strongest associations are those occurring between the pairs pineapple and banana ($s = 0.40$), camphor and minty ($s = 0.33$), apricot and peachy ($s = 0.33$), tarry and smokey ($s = 0.31$) and lily and narcissus ($s = 0.30$). These results show that the notes used in perfumery are generally rather independent.

The fine analysis of associations between notes by clustering, using the complete linkage method, allowed us to build the dendrogram presented in Figure 1. If an association is considered as being significant only when the corresponding similarity coefficient is above the threshold of 0.08 on the dendrogram, 14 notes are not part of groups. Among these notes, some present small similarity coefficients with other notes or groups of notes but others, such as tea, almond and peppery, are not associated at all. The remaining 60 notes are distributed between 27 groups comprising two to four notes as indicated in the dendrogram.

The existence of 14 isolated notes and of 27 groups containing two to four notes leads us to conclude that there cannot exist a small number of primary odors in the system of description of odors used by Arctander, and furthermore that the groups obtained cannot be organized according to a strict hierarchy. Thus our analysis confirms that the terms employed to describe odors are generally independent, as was to be expected from such a long and carefully tested description system. This independence is verified further by a factorial analysis performed on the similarity matrix for 24 notes, which shows that the first principal components represent a very small fraction of the total variance (Chastrette *et al.*, 1986a), thus limiting the usefulness of the representation of notes in the first factorial planes.

Should one then turn to the concept of a continuum in which all associations among notes are more or less arbitrary? We found that this is not completely the case and in the following discussion we intend to examine some of the observed associations in order to try to interpret them from the two points of view of their agreement with the experience of perfumers and of the possibility of establishing structure−odor relationships.

Isolated notes

The so-called isolated notes are notes for which one does not observe any association at a level of ≥ 0.08 in the similarity matrix.

The almond note presents no notable resemblance (all the similarity coefficients are

<0.06). An examination of the structure of almondy products shows that they often comprise a benzene nucleus carrying an aldehyde or nitrile group.

The sulfuraceous note presents practically no similarity: among its 26 citations this note is associated only five times with the herbaceous note, four times with the green note and four times with the oily note. It has been pointed out (Jaubert, 1983) that the products defined by this note contain a sulfur atom and we found it to be true for the population studied.

Almost isolated notes

These notes possess one or several coefficients the value of which is >0.08 in the similarity matrix. This phenomenon is encountered when a note presents high similarity coefficients with some notes which otherwise do not have any marked similarity between them. Since the aggregation tends to be made among the most similar notes, the less similar notes appear almost isolated on the dendrogram.

The anisic note, encountered 37 times in the data bank, is only very weakly associated with herbaceous ($s = 0.09$) and floral ($s = 0.08$) notes. Examination of products presenting this note shows that they correspond to well-defined chemical structures (work in progress).

Sandalwood presents a marked similarity only with medicinal ($s = 0.12$), woody ($s = 0.10$) and amber ($s = 0.10$). The similarity with the woody note is weaker than would be expected; this was also noted by Arctander (1969). It has been recently shown (Shankaranarayana *et al.*, 1984; Witteven *et al.*, 1987) that the sandalwood note lends itself very well to the establishment of structure–odor relations, thus making possible the synthesis of new products possessing this odor.

The note caramellic appears 64 times in the sample and is associated with fruity, buttery and spicy notes. In a recent review, Ohloff (1986) points out that caramellic products usually possess the system alkyl/enol/carbonyl in their chemical structure.

Paired notes

Several classes are constituted of two notes exhibiting a level of aggregation falling within the extremum values of 0.09 for the lilac–jasmin pair and 0.4 for the pineapple–banana pair. The following pairs look interesting from the viewpoint of structure–odor relationships.

The amber note presents no notable similarity coefficient except with the note musty ($s = 0.13$). This note has been thoroughly studied by Ohloff and co-workers (Ohloff *et al.*, 1985; Ohloff, 1986), who showed that it corresponds to a well-defined geometry of the molecule.

The musky note presents a pronounced similarity coefficient only with the animal note ($s = 0.21$). Several workers have shown that this note depends very directly on the structure of molecules inside a particular family (Brugger and Jurs, 1977; Chastrette *et al.*, 1986b; Narvaez *et al.*, 1986). The association with the animal note is probably due to the animal origin of these products.

The notes of the pair camphor–minty present a pronounced similarity coefficient and are associated, although to a lesser extent, with the woody and piney notes when one considers the average linkage method. Our work on structure–odor relationships for camphoraceous substances (Chastrette, 1981; Eminet and Chastrette, 1983) and for

minty substances (work in progress) shows that the two groups have closely related structures.

Other groups

The meaning of the group fruity – floral – green has to be related either to high numbers of associations or to similarities in the chemical structure.

The fruity note is associated with 28 out of the other 73 notes and appears 715 times in the description of the 2467 odorants of the sample. When examining the corresponding chemical structures it can be seen that they possess a considerable chemical variety and that it would be illusory to attempt to attribute to this note a particular type of chemical structure. An explanation of these associations is to be found in the domain of semantics. The fruity note can be considered as a generic note, only implying that a particular note belongs to the family of fruity notes. Thereafter the odor has to be more precisely identified by actually naming the fruit in question. To support this view we found that the note fruity is cited alone in only 2.6% of its occurrences, while associated with odors of specific fruits is cited in 24% of the cases.

The floral note is associated with 23 of the 73 other notes and appears 678 times in the sample. We can say that floral is a generic term eventually qualified by the denomination of the exact flower in question. This note is cited alone in only 5.4% of its occurrences and is associated with the name of a specific flower (jasmin, orange-blossom, hyacinth, lily, lilac, rose) in 17.2% of its citations.

The green note is associated with 20 other notes and appears 558 times in the sample. The analysis of co-occurrences shows that it is closely associated with fruity, floral, herbaceous, oily, woody and rose. The corresponding chemical structures possess a large variety and it is likely that structure – odor relationships studies in this domain would prove less than fruitful.

The ethereal – apple – pineapple – pear group, which is made of fruity notes, appears homogeneous and consistent with the experience of perfumers. An examination of the chemical structure of different products presenting this note shows they often contain ester groups.

In the oily – fatty – waxy group, the oily note appears 306 times, associated with 14 other notes. It presents marked similarities with floral, herbaceous and fatty. The fatty note appears 123 times and is associated with nine notes and especially with the oily note. As for the waxy note, it appears 72 times in the data bank and presents a marked similarity with the fatty note. An examination of the chemical structure of products presenting these notes shows that they generally contain a long hydrocarbon chain.

Conclusion

The similarities existing among the 74 notes used in the description of the odors of 2467 substances have been evidenced by analysis of co-occurrences, calculation of the similarity coefficients of Ochiai and subsequent cluster analysis.

As a rule the similarity between notes is rather weak and shows the relative independence of the descriptors used by Arctander.

The hierarchy of notes shown by the dendrogram is weak. Significant groups do not exceed five or six notes. The olfactive universe is thus better described by a model based on a continuum, already proposed by Holley and MacLeod (1977), than by a

strongly hierarchic model. This implies that a small number of primary odors cannot be considered.

For certain isolated or almost isolated notes, it is possible to associate odors with a structural particularity such as the presence of a sulfur atom for sulfuraceous notes or a particular geometric configuration of the molecule for the amber note. These notes lend themselves well to structure–odor relationships. The associations by pairs are often well explained either by chemical structure similarity (camphoraceous/minty, oily/fatty) or by semantic associations (musk/animal, amber/musty or acid/sour).

A number of associations between notes, although strong (e.g. floral, fruity), are not due to structural resemblances, but seem to be related to the mental process of odor descriptions by successive approximations. For example, a note described as fruity is further qualified by the name of the fruit. The groups thus created are actually associated by their common vegetable origin rather than by similarities in their chemical structures.

The conclusions of this study, drawn from descriptions of odors by a single perfumer, cannot yet be extended to the olfactive universe of perfumery in general. For such an extension it will be necessary to study the description systems used by other perfumers or groups of perfumers.

Finally, the generalization of these findings to the human olfactory universe in the largest sense of the term should rely on a larger study, taking into account several notes that have not been very deeply considered in perfumery, for example unpleasant odors.

Acknowledgements

We thank Mrs T.Roudnitska, Mr E.Roudnitska and Professor A.Holley for their help during this work and the Fondation de France for financial support.

References

Arctander,S. (1969) *Perfume and Flavor Chemicals.* Monclair, NJ.

Benzecri,J.P. (1980) *L'analyse des données. I. La taxinomie.* Dunod, Paris.

Boelens,H. (1974) Relationship between the chemical structure of compounds and their olfactive properties. *Cosmet. Perfum.*, **89**, 70–78.

Boelens,H. (1983) Structure–activity relationships in chemoreception by human olfaction. *Trends Pharm. Sci.*, **4**, 421–426.

Brud,W.S. (1986) Words versus odours: how perfumers communicate. *Perfum. Flav.*, **11**, 27–44.

Brugger,W.E. and Jurs,P.C. (1977) Extraction of important molecular features of musk compounds using pattern recognition techniques. *J. Agric. Food Chem.*, **5**, 1158–1164.

Cerbelaud,R. (1951) *Formulaire des parfumeurs.* Eyrolles, Paris.

Chastrette,M. (1981) An approach to a classification of odours using physicochemical parameters. *Chem. Senses*, **6**, 157–163.

Chastrette,M., Elmouaffek,A. and Zakarya,D. (1986a) Etude statistique multidimensionnelle des similarités entre 24 notes utilisées en parfumerie. *C.R. Acad. Sci. Paris*, **303**, Série II, 1209–1214.

Chastrette,M., Zakarya,D. and Elmouaffek,A. (1986b) Relations structure–odeur dans la famille des muscs benzéniques nitrés. *Eur. J. Med. Chem.*, **21**, 505–510.

Cole,L.C. (1949) The measurement of interspecific association. *Ecology*, **30**, 411–424.

Crocker,E.C. and Henderson,F.L. (1927) Analysis and classification of odors: an effort to develop a workable method. *Am. Perfumer Ess. Oil Rev.*, **22**, 325–356.

Dagnelie,P. (1984) *Analyse Statistique à plusieurs variables.* Presses Agronomiques de Gembloux, Gembloux.

Dore,V., Gordon,G. and Jaubert,J.N. (1984) Approche factorielle des relations entre structure chimique et note odorante. *C.R. Acad. Sci. Paris*, **299**, Série II, 315–320.

Dravnieks,A. (1982) Odor quality: semantically generated multidimensional profiles are stable. *Science*, **218**, 799–801 (and references cited therein).

Eminet,B.P. and Chastrette,M. (1983) Discrimination of camphoraceous substances using physico-chemical parameters. *Chem. Senses*, 7,293–300.

Harper,R. (1975) Some chemicals representing particular odour qualities. *Chem. Senses Flavor*, **1**, 353–357.

Holley,A. and MacLeod,P. (1977) Transduction et codage des informations olfactives chez les vertébrés. *J. Physiol.*, **73**, 725–828.

Iglesias,A. (1975) Contribution à l'analyse des tableaux de distances et de similarité: Applications à la Biologie. Thèse 3ème Cycle, Lyon.

Jaubert,J.N. (1983) Analyse multidimensionnelle de 1400 molécules utilisées comme arômes alimentaires, constitution d'une base de données informatiques et traitement factoriel de quelques corrélations significatives. Thèse de docteur-ingénieur, Toulouse.

Jaubert,J.N., Gordon,G. and Dore,J.C. (1986) Classification of odours and their sensorial perception. *Quintessenza*, **5**, 27–42.

Kastner,D. (1973) Die Beschreibung und Klassifizierung von Gerüchen. *Parfüm. Kosmet.*, **54**, 97–106.

Massart,D.L. and Kaufman,L. (1983) *The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis*. Wiley, New York.

Narvaez,J.N., Lavine,B.K. and Jurs,P.C. (1986) Structure–activity studies of musk odorants using pattern recognition: bicyclo- and tricyclo-benzoids. *Chem. Senses*, **11**, 145–156.

Oichiai,A. (1957) Zoogeographic studies on the soleoid fishes found in Japan and its neighbouring regions. *Bull. Jap. Sci. Fish.*, **22**, 526–530.

Ohloff,G. (1986) Chemistry of odor stimuli. *Experientia*, **42**, 271–279.

Ohloff,G., Giersh,W., Pickenhagen,W., Funer,A. and Frei,B. (1985) Significance of the geminal dimethyl group in the odor principle of Ambrox. *Helv. Chim. Acta*, **68**, 2022–2029.

Randebrock,R.E. (1985) Geruch und Konstitution: Eine Beweisfuhrung fur die Molekulartheorie des Geruches. *Parfüm. Kosmet.*, **66**, 435–443.

Roux,M. (1985) *Algorithmes de classification*. Masson, Paris.

Shankaranarayana,K.H. and Parthasarathi,K. (1984) Synthetic sandalwood aroma chemicals. *Perfum. Flav.*, **9**, 17–20.

Sokal,R.R. and Sneath,P. (1963) *Principles of Numerical Taxonomy*. Freeman, San Francisco.

Witteven,J.G. and Van der Weerdt,A.J.A. (1987) Structure–odour relationships of some new synthetic sandalwood aroma chemicals. *Rec. Trav. Chim. Pays-Bas*, **106**, 29–34.

Received on October 21, 1987; accepted on March 4, 1988